Distributional robustness of K-class estimators and the PULSE

MARTIN EMIL JAKOBSEN AND JONAS PETERS

Presenter: Will Zhang

University of Washington

March 11, 2023

Table of contents

Introduction

Problem: Casual Models are not prediction optimal under bounded perturbations

This Paper:

- Connect K-class estimator and anchor regression
- propose p-uncorrelated least squares estimator (PULSE)

Distributionally robust optimization

When considering distributional robustness, we are interested in finding a γ that minimizes the worst case expected squared prediction error over a class of distributions, $\mathcal{F}$:

$$\arg\min_{\gamma} \sup_{F \in \mathcal{F}} E_F \left[\left(Y - \gamma^\top X \right)^2 \right].$$

The true causal coefficient γ_0 minimizes the equation when $\mathcal{F}$ is the set of all possible interventions on X. The OLS solution is optimal when $\mathcal{F}$ only contains the training distribution. OLS solution and the true causal coefficient constitute the end points of a spectrum of estimators that are prediction optimal under a certain class of distributions.

Structural equation model setup

Consider a possibly cyclic linear SEM:

$$\begin{bmatrix} Y & X^\top & H^\top \end{bmatrix} := \begin{bmatrix} Y & X^\top & H^\top \end{bmatrix} B + A^\top M + \varepsilon^\top,$$

$Y \in \mathbb{R}$ is the endogenous target, $X \in \mathbb{R}^d$ are the observed endogenous variables, $H \in \mathbb{R}^r$ are hidden endogenous variables, and $A \in \mathbb{R}^q$ are exogenous variables independent from the unobserved noise innovations ε. Under intervention do$(A := v)$:

$$\begin{bmatrix} Y_v & X_v^\top & H_v^\top \end{bmatrix} := \begin{bmatrix} Y_v & X_v^\top & H_v^\top \end{bmatrix} B + v^\top M + \varepsilon^\top.$$

Let $(\mathbf{Y}, \mathbf{X}, \mathbf{H}, \mathbf{A})$ consist of n row-wise random vector (Y, X, H, A) and consider the single equation of interest

$$\mathbf{Y} = \mathbf{X}\gamma_0 + \mathbf{A}\beta_0 + \mathbf{H}\eta_0 + \boldsymbol{\varepsilon}_Y = \mathbf{X}\gamma_0 + \mathbf{A}\beta_0 + \tilde{\mathbf{U}}_Y$$

K-class estimator & Anchor Regression

Using nonsample information, estimate nonzero coefficients $\mathbf{Z}_* \subset [\mathbf{XA}]$

$$\hat{\alpha}^n_{\mathrm{K}}(\kappa) = \left(\mathbf{Z}_*^\top \left(I - \kappa P_{\mathbf{A}}^\perp\right) \mathbf{Z}_*\right)^{-1} \mathbf{Z}_*^\top \left(I - \kappa P_{\mathbf{A}}^\perp\right) \mathbf{Y},$$

$P_{\mathbf{A}}^\perp$ is the projection onto the orthogonal complement of the column space of $\mathbf{A}$. For a fixed $\kappa \in [0,1)$, K-class estimators can be represented by Anchor Regression

$$\hat{\alpha}^n_{\mathrm{K}}(\kappa) = \arg\min_{\alpha} l^n_{\mathrm{OLS}}(\alpha) + \kappa/(1-\kappa) l^n_{\mathrm{IV}}(\alpha)$$

where l^n_{OLS} and l^n_{IV} are the empirical OLS and TSLS loss functions. Above allows, for a fixed κ, regardless of identifiability

$$\hat{\alpha}^n_{\mathrm{K}}(\kappa) \xrightarrow[n\to\infty]{P} \arg\min_{\alpha} \sup_{v \in C(\kappa)} E^{\mathrm{do}(A:=v)}\left[\left(Y - \alpha^\top Z_*\right)^2\right],$$

where $C(\kappa) := \left\{v : \Omega \to \mathbb{R}^q : \mathrm{Cov}(v, \varepsilon) = 0, E\left[vv^\top\right] \preceq \frac{1}{1-\kappa} E\left[AA^\top\right]\right\}.$

The PULSE estimator

PULSE as a data-driven K-class estimator

$$\hat{\alpha}^n_{\mathrm{K}}\left(\lambda^\star_n / \left(1+\lambda^\star_n\right)\right) = \arg\min_{\alpha} l^n_{\mathrm{OLS}}(\alpha) + \lambda^\star_n l^n_{\mathrm{IV}}(\alpha),$$

where

$$\lambda^\star_n := \inf\left\{\lambda > 0 : \begin{array}{c}\text{testing } \operatorname{Corr}\left(A, Y - Z\hat{\alpha}^n_{\mathrm{K}}(\lambda/(1+\lambda))\right) = 0 \\ \text{yields a } p\text{-value } \geq p_{\min}\end{array}\right\},$$

for some pre-specified level of the hypothesis test $p_{\min} \in (0,1)$.
As a constrained optimization problem, define primal problems:

$$\hat{\alpha}^n_{\mathrm{Pr}}(t) := \begin{array}{ll}\arg\min_{\alpha} & l^n_{\mathrm{OLS}}(\alpha) \\ \text{subject to} & l^n_{\mathrm{IV}}(\alpha) \leq t.\end{array}$$

$$t^\star_n := \sup\left\{t : \text{testing } \operatorname{Corr}\left(A, Y - Z\hat{\alpha}^n_{P_{\mathrm{r}}}(t)\right) = 0 \text{ yields a } p\text{-value } \geq p_{\min}\right\}$$

$$\hat{\alpha}^n_{\mathrm{K}}\left(\lambda^\star_n / \left(1+\lambda^\star_n\right)\right) = \hat{\alpha}^n_{\mathrm{Pr}^n}\left(t^\star_n\right)$$

The PULSE estimator

Third, PULSE can be written as

$$\begin{aligned} \arg\min_{\alpha} \quad & l^n_{\mathrm{OLS}}(\alpha; \mathbf{Y}, \mathbf{Z}) \\ \text{subject to} \quad & \alpha \in \mathcal{A}_n\left(1 - p_{\min}\right), \end{aligned}$$

where $\mathcal{A}_n\left(1 - p_{\min}\right)$ is the nonconvex acceptance region for our test of uncorrelatedness.

Interpretation of estimator:

- among all coefficients yielding uncorrelatedness, choose the one that yields the best prediction.

Setup and assumptions

Assume the SEM.
$[\mathbf{YXH}]\Gamma = \mathbf{A}M + \boldsymbol{\varepsilon}$ and $[\mathbf{YXH}] = \mathbf{A}\Pi + \varepsilon\Gamma^{-1}$, where $\Gamma := I - B$ and $\Pi := M\Gamma^{-1}$. Assume Γ has a unity diagonal, such that the target equation of interest is given by

$$\mathbf{Y} = \mathbf{X}\gamma_0 + \mathbf{A}\beta_0 + \mathbf{H}\eta_0 + \boldsymbol{\varepsilon}_Y = \mathbf{Z}\alpha_0 + \tilde{\mathbf{U}}_Y,$$

where $(1, -\gamma_0, -\eta_0) \in \mathbb{R}^{(1+d+r)}, \beta_0 \in \mathbb{R}^q$ and $\boldsymbol{\varepsilon}_Y$ are the first columns of Γ, M and $\boldsymbol{\varepsilon} \in \mathbb{R}^{n\times d}$ respectively, $\mathbf{Z} := [\mathbf{X}\ \mathbf{A}], \alpha_0 = (\gamma_0, \beta_0) \in \mathbb{R}^{d+q}$ and $\tilde{\mathbf{U}}_Y := \mathbf{H}\eta_0 + \boldsymbol{\varepsilon}_Y$.

ASSUMPTION 2.1 (GLOBAL ASSUMPTIONS)

(a) (Y, X, H, A) is generated in accordance with the SEM in equation (2.1); (b) $\rho(B) < 1$ where $\rho(B)$ is the spectral radius of B; (c) ε has jointly independent marginals $\varepsilon_1, \ldots, \varepsilon_{d+1+r}; (d) A$ and ε are independent;; (e) No variable in Y, X and H is an ancestor of A, that is, A is exogenous; $(f) E\left[\|\varepsilon\|_2^2\right], E\left[\|A\|_2^2\right] < \infty; (g) E[\varepsilon] = 0$; (h) $\mathrm{Var}(A) \succ 0$, i.e., the variance matrix of A is positive definite; (i) $\mathbf{A}^\top \mathbf{A}$ is almost surely of full rank;

ASSUMPTION 2.2 (FINITE SAMPLE ASSUMPTIONS)

(a) $\mathbf{Z}_*^\top \mathbf{Z}_*$ is almost surely of full rank; (b) $\mathbf{A}^\top \mathbf{Z}_*$ is almost surely of full column rank. (c) $\mathbf{X}^\top \mathbf{X}$ is almost surely of full rank;

ASSUMPTION 2.3 (POPULATION ASSUMPTIONS)

(a) $\mathrm{Var}(Z_*) \succ 0$, i.e., the variance matrix of Z_* is positive definite; $(b) E\left[\mathbf{A}\mathbf{Z}_*^\top\right]$ is of full column rank.

Setup of anchor regression

$P_{\mathbf{A}} = \mathbf{A}\left(\mathbf{A}^\top \mathbf{A}\right)^{-1}\mathbf{A}^\top$ is the orthogonal projection onto the column space of $\mathbf{A}$. Define loss functions and their sample analog:

$$l_{\mathrm{OLS}}(\gamma; Y, X) := E\left(Y - \gamma^\top X\right)^2,$$

$$l_{\mathrm{IV}}(\gamma; Y, X, A) := E\left(A\left(Y - \gamma^\top X\right)\right)^\top E\left(AA^\top\right)^{-1} E\left(A\left(Y - \gamma^\top X\right)\right),$$

$$l^n_{\mathrm{OLS}}(\gamma; \mathbf{Y}, \mathbf{X}) := n^{-1}(\mathbf{Y} - \mathbf{X}\gamma)^\top(\mathbf{Y} - \mathbf{X}\gamma),$$

$$l^n_{\mathrm{IV}}(\gamma; \mathbf{Y}, \mathbf{X}, \mathbf{A}) := n^{-1}(\mathbf{Y} - \mathbf{X}\gamma)^\top P_{\mathbf{A}}(\mathbf{Y} - \mathbf{X}\gamma),$$

With penalty parameter $\lambda > -1$

$$\gamma_{\mathrm{AR}}(\lambda) := \arg\min_{\gamma \in \mathbb{R}^d} \left\{l_{\mathrm{OLS}}(\gamma) + \lambda l_{\mathrm{IV}}(\gamma)\right\},$$

$$\hat{\gamma}^n_{\mathrm{AR}}(\lambda) := \arg\min_{\gamma \in \mathbb{R}^d} \left\{l^n_{\mathrm{OLS}}(\gamma) + \lambda l^n_{\mathrm{IV}}(\gamma)\right\}.$$

Distributional robustness of AR

The solution to AR empirical minimization problem can be written as

$$\hat{\gamma}^n_{\mathrm{AR}}(\lambda) = \left[\mathbf{X}^\top (I + \lambda P_{\mathbf{A}})\, \mathbf{X}\right]^{-1} \mathbf{X}^\top (I + \lambda P_{\mathbf{A}})\, \mathbf{Y},$$

AR is well identified even if the model is under-identified. AR aim to possess a interventional robustness instead of inferring causality.

$$\gamma_{\mathrm{AR}}(\lambda) = \underset{\gamma \in \mathbb{R}^d}{\arg\min} \sup_{v \in C(\lambda)} E^{\mathrm{do}(A:=v)}\left[\left(Y - \gamma^\top X\right)^2\right],$$

where
$C(\lambda) := \left\{v : \Omega \to \mathbb{R}^q : \mathrm{Cov}(v, \varepsilon) = 0, E\left(vv^\top\right) \preceq (\lambda + 1) E\left(AA^\top\right)\right\}.$

Setup of K-class estimators

Given nonsample information about which components of γ_0 and β_0 are zero, partition $\mathbf{X} = [\mathbf{X}_* \mathbf{X}_{-*}] \in \mathbb{R}^{n\times(d_1+d_2)}$, $\mathbf{A} = [\mathbf{A}_* \mathbf{A}_{-*}] \in \mathbb{R}^{n\times(q_1+q_2)}$ and $\mathbf{Z} = [\mathbf{Z}_* \mathbf{Z}_{-*}] = [\mathbf{X}_* \mathbf{A}_* \mathbf{X}_{-*} \mathbf{A}_{-*}]$ with $\mathbf{Z} \in \mathbb{R}^{n\times((d_1+q_1)+(d_2+q_2))}$, where $\mathbf{X}_{-*}$ and $\mathbf{A}_{-*}$ correspond to the components of γ_0 and β_0 are zero, respectively. Similarly, $\gamma_0 = (\gamma_{0,*}, \gamma_{0,-*})$, $\beta_0 = (\beta_{0,*}, \beta_{0,-*})$ and $\alpha_0 = (\alpha_{0,*}, \alpha_{0,-*}) = (\gamma_{0,*}, \beta_{0,*}, \gamma_{0,-*}, \beta_{0,-*})$. Interested equation:

$$\mathbf{Y} = \mathbf{X}_*\gamma_{0,*} + \mathbf{X}_{-*}\gamma_{0,-*} + \mathbf{A}_*\beta_{0,*} + \mathbf{A}_{-*}\beta_{0,-*} + \tilde{\mathbf{U}}_Y = \mathbf{Z}_*\alpha_{0,*} + \mathbf{U}_Y$$

where $\mathbf{U}_Y = \mathbf{X}_{-*}\gamma_{0,-*} + \mathbf{A}_{-*}\beta_{0,-*} + \mathbf{H}\eta_0 + \boldsymbol{\varepsilon}_Y$. When well-defined, with parameter $\kappa \in \mathbb{R}$ for a simultaneous estimation of $\alpha_{0,*}$:

$$\hat{\alpha}^n_{\mathrm{K}}(\kappa; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}) = \left(\mathbf{Z}_*^\top \left(I - \kappa P_{\mathbf{A}}^\perp\right) \mathbf{Z}_*\right)^{-1} \mathbf{Z}_*^\top \left(I - \kappa P_{\mathbf{A}}^\perp\right) \mathbf{Y},$$

where $I - \kappa P_{\mathbf{A}}^\perp = I - \kappa(I - P_{\mathbf{A}}) = (1-\kappa)I + \kappa P_{\mathbf{A}}$.

Distributional robustness of k-class

K-class estimator with no included exogenous variables, for $\kappa < 1$ coincides with the AR estimator with penalty parameter $\lambda = \kappa/(1-\kappa)$, i.e., $\hat{\gamma}^n_{\mathrm{K}}(\kappa) = \hat{\gamma}^n_{\mathrm{AR}}\left(\frac{\kappa}{1-\kappa}\right)$, or Equivalently $\hat{\gamma}^n_{\mathrm{AR}}(\lambda) = \hat{\gamma}^n_{\mathrm{K}}(\lambda/(1+\lambda))$ for any $\lambda > -1$.

K-class estimator, for a fixed $\kappa < 1$, inherits distributional robustness property:

$$\hat{\gamma}^n_{\mathrm{K}}(\kappa) \xrightarrow[n\to\infty]{P} \arg\min_{\gamma\in\mathbb{R}^d} \sup_{v\in C(\kappa/(1-\kappa))} E^{\mathrm{do}(A:=v)}\left[\left(Y - \gamma^\top X\right)^2\right],$$

where

$C(\kappa/(1-\kappa)) = \left\{v : \Omega \to \mathbb{R}^q : \mathrm{Cov}(v,\varepsilon) = 0, E\left[vv^\top\right] \preceq \frac{1}{1-\kappa} E\left[AA^\top\right]\right\}$.

The K-class estimators as penalized regression estimators

For TSLS loss function, regress $\mathbf{Y}$ on the included endogenous and exogenous variables $\mathbf{Z}_*$ using the exogeneity of $\mathbf{A}$ and $\mathbf{A}_{-*}$ as instruments and for the OLS loss function, regress $\mathbf{Y}$ on $\mathbf{Z}_*$. We define

$$l_{\mathrm{K}}(\alpha; \kappa, Y, Z_*, A) = (1-\kappa) l_{\mathrm{OLS}}(\alpha; Y, Z_*) + \kappa l_{\mathrm{IV}}(\alpha; Y, Z_*, A)$$
$$l_{\mathrm{K}}^n(\alpha; \kappa, \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}) = (1-\kappa) l_{\mathrm{OLS}}^n(\alpha; \mathbf{Y}, \mathbf{Z}_*) + \kappa l_{\mathrm{IV}}^n(\alpha; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A})$$

Proposition 2.1. Consider one of the following scenarios: 1) $\kappa < 1$ and 2.2 (a) holds, or 2) $\kappa = 1$ and 2.2(b) holds. The estimator minimizing the empirical loss function is almost surely well-defined and coincides with the K-class estimator. That is, it almost surely holds that

$$\hat{\alpha}_{\mathrm{K}}^n(\kappa; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}) = \underset{\alpha \in \mathbb{R}^{d_1+q_1}}{\arg\min}\, l_{\mathrm{K}}^n(\alpha; \kappa, \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}).$$

The K-class estimators as penalized regression estimators

Assuming $\kappa \neq 1$, we can rewrite above equation to

$$\hat{\alpha}_{\mathrm{K}}^{n}(\kappa; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}) = \underset{\alpha \in \mathbb{R}^{d_1+q_1}}{\arg\min} \left\{ l_{\mathrm{OLS}}^{n}(\alpha; \mathbf{Y}, \mathbf{Z}_*) + \frac{\kappa}{1-\kappa} l_{\mathrm{IV}}^{n}(\alpha; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}) \right\}.$$

Proposition 2.2. Consider one of the following scenarios: 1) $\kappa \in [0, 1\ )$ and 2.3 (a) holds, or 2) $\kappa = 1$ and 2.3 (b) holds. It holds that $(\hat{\alpha}_{K}^{n}(\kappa; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}))_{n \geq 1}$ is an asymptotically welldefined sequence of estimators. Furthermore, the sequence consistently estimates the well-defined population K-class estimand. That is,

$$\hat{\alpha}_{\mathrm{K}}^{n}(\kappa; \mathbf{Y}, \mathbf{Z}_*, \mathbf{A}) \xrightarrow[n \to \infty]{P} \alpha_{\mathrm{K}}(\kappa; Y, Z_*, A) := \underset{\alpha \in \mathbb{R}^{d_1+q_1}}{\arg\min}\, l_{\mathrm{K}}(\alpha; \kappa, Y, Z_*, A).$$

Distributional robustness of general K-class estimators

K-class estimator is prediction optimal under a set of interventions on all exogenous A up to a certain strength.

THEOREM 2.1. Let Assumption 2.1 hold. For any fixed $\kappa \in [0,1)$ and $Z_* = (X_*, A_*)$ with $X_* \subseteq X$ and $A_* \subseteq A$, we have, whenever the population K-class estimand is well-defined, that

$$\alpha_{\mathrm{K}}(\kappa; Y, Z_*, A) = \underset{\alpha \in \mathbb{R}^{d_1+q_1}}{\arg\min} \sup_{v \in C(\kappa)} E^{\mathrm{do}(A:=v)}\left[\left(Y - \alpha^\top Z_*\right)^2\right],$$

where $C(\kappa) := \left\{v : \Omega \to \mathbb{R}^q : \mathrm{Cov}(v, \varepsilon) = 0, E\left[vv^\top\right] \preceq \frac{1}{1-\kappa} E\left[AA^\top\right]\right\}$.
$E^{\mathrm{do}(A:=v)}$ denotes the expectation with respect to the distribution entailed under the intervention $\mathrm{do}(A := v)$.

Objective function

Predict target Y from endogenous and possibly exogenous regressors Z.

$$\begin{aligned} \hat{\alpha}^n_{PULSE}(p_{min}) = & \operatorname{argmin}_\alpha \ell^n_{OLS}(\alpha; \mathbf{Y}, \mathbf{Z}) \\ & \text{s.t. p-value}(T(\alpha; \mathbf{Y}, \mathbf{Z}, \mathbf{A})) \geq p_{min}, \end{aligned} \tag{3.1}$$

where $T(\alpha; \mathbf{Y}, \mathbf{Z}, \mathbf{A})$ is finite sample test statistic for $\mathcal{H}_0(\alpha)$: $\text{Corr}\left(A, Y - \alpha^\top Z\right) = 0$, and p_{min} is a pre-specified level of the hypothesis test.

- choose the solution that is 'closest' to the OLS solution while maintaining uncorrelatedness
- a K-class estimator with a data- driven κ

Setup and assumptions

Assume the SEM, the structural equation of interest is

$$Y = \gamma_0^\top X + \eta_0^\top H + \beta_0^\top A + \varepsilon_Y$$

Assume that we have some nonsample information: $d_2 = d - d_1$ and $q_2 = q - q_1$ coefficients of γ_0 and β_0, respectively, are zero. Now write

- $Z = \left[X_*^\top A_*^\top\right]^\top \in \mathbb{R}^{d_1+q_1}, \mathbf{Z} = [\mathbf{X}_* \mathbf{A}_*] \in \mathbb{R}^{n\times(d_1+q_1)}$
- $\alpha_0 := \left(\gamma_{0,*}^\top, \beta_{0,*}^\top\right)^\top :\in \mathbb{R}^{d_1+q_1}$.

That is, $Y = \alpha_0^\top Z + U_Y$, where $U_Y = \eta_0^\top H + \varepsilon_Y$.
Global assumptions of Assumption 2.1 hold

Assumption 3.1

(a) $A \perp\!\!\!\perp U_Y$; (b) $E[A] = 0$.

Assumption 3.2

ε has nondegenerate marginals.

Assumption 3.3

(a) $\mathbf{Z}^\top \mathbf{Z}$ is of full rank; $(b)\mathbf{A}^\top \mathbf{Z}$ is of full rank.

Assumption 3.4

$[\mathbf{Z}\ \mathbf{Y}]$ is of full column rank.

Assumption 3.5

$E\left[AZ^\top\right]$ is of full rank.

Testing for vanishing correlation

Define $T_n^c : \mathbb{R}^{d_1+q_1} \to \mathbb{R}$ by

$$T_n^c(\alpha) := c(n)\frac{l_{\mathrm{IV}}^n(\alpha)}{l_{\mathrm{OLS}}^n(\alpha)} = c(n)\frac{\|P_{\mathbf{A}}(\mathbf{Y} - \mathbf{Z}\alpha)\|_2^2}{\|\mathbf{Y} - \mathbf{Z}\alpha\|_2^2},$$

where $c(n)$ is a function that will typically scale linearly in n. Denote the $1 - p$ quantile of the central Chi-Squared distribution with q degrees of freedom by $Q_{\chi_q^2}(1 - p)$. By standard limiting theory we can test $\mathcal{H}_0(\alpha)$ in the following manner

LEMMA 3.1 (LEVEL AND POWER OF THE TEST)

Let Assumptions 3.1, 3.2, and 3.4 hold and assume that $c(n) \sim n$ as $n \to \infty$. For any $p \in (0, 1)$ and any fixed α, the statistical test rejecting the null hypothesis $\mathcal{H}_0(\alpha)$, if $T_n^c(\alpha) > Q_{\chi_q^2}(1 - p)$, has point-wise asymptotic level p and point-wise asymptotic power of 1 against all alternatives as $n \to \infty$.

The PULSE estimator

For simplicity, $c(n) = n$ and define the acceptance region with level $p_{\min} \in (0,1)$ as $\mathcal{A}_n\left(1 - p_{\min}\right) := \left\{\alpha \in \mathbb{R}^{d_1+q_1} : T_n(\alpha) \leq Q_{\chi^2_q}(1- \ \ p_{\min})\right\}$. Consider

$$\hat{\alpha}^n_{\mathrm{PULSE}}\left(p_{\min}\right) := \begin{array}{ll} \arg\min_\alpha & l^n_{\mathrm{OLS}}(\alpha) \\ \text{subject to} & T_n(\alpha) \leq Q_{\chi^2_q}\left(1 - p_{\min}\right). \end{array}$$

- Generally a nonconvex optimization problem
- Objective function has a unique solution that coincides with the solution of a strictly convex, quadratically constrained quadratic program (QCQP) with a data-dependent constraint bound.

Primal representation of PULSE

Derive a QCQP representation of PULSE (primal PULSE). For all $t \geq 0$ define the empirical primal minimization (Primal.t.n) problem

$$\begin{array}{ll} \text{minimize}_\alpha & l^n_{\mathrm{OLS}}(\alpha; \mathbf{Y}, \mathbf{Z}) \\ \text{subject to} & l^n_{\mathrm{IV}}(\alpha; \mathbf{Y}, \mathbf{Z}, \mathbf{A}) \leq t \end{array}$$

Under suitable assumptions these problems are solvable

LEMMA 3.2 (UNIQUE SOLVABILITY OF THE PRIMAL)

Let Assumption 3.3 hold. It holds that $\alpha \mapsto l^n_{\mathrm{OLS}}(\alpha)$ and $\alpha \mapsto l^n_{\mathrm{IV}}(\alpha)$ are strictly convex and convex, respectively. Furthermore, for any $t > \inf_\alpha l^n_{\mathrm{IV}}(\alpha)$ it holds that the constrained minimization problem (Primal.t.n) has a unique solution and satisfies Slater's condition. In the under- and just-identified setup the constraint bound requirement is equivalent to $t > 0$ and in the over-identified setup to $t > l^n_{\mathrm{IV}}(\hat{\alpha}^n_{\mathrm{TSLS}})$, where $\hat{\alpha}^n_{\mathsf{TSLS}} = \left(\mathbf{Z}^\top P_{\mathbf{A}} \mathbf{Z}\right)^{-1} \mathbf{Z}^\top P_{\mathbf{A}} \mathbf{Y}$

PULSE and the primal PULSE

Restrict the constraint bounds to $D_{\mathrm{Pr}} := (\inf_{\alpha} l^n_{\mathrm{IV}}(\alpha), l^n_{\mathrm{IV}}(\hat{\alpha}^n_{\mathrm{OLS}})]$
Define $t^\star_n(p_{\min})$ as the data-dependent constraint bound given by

$$t^\star_n(p_{\min}) := \sup\left\{ t \in \left(\inf_{\alpha} l^n_{\mathrm{IV}}(\alpha), l^n_{\mathrm{IV}}(\hat{\alpha}^n_{\mathrm{OLS}})\right] : T_n(\hat{\alpha}^n_{\mathrm{Pr}}(t)) \leq Q_{\chi^2_q}(1-p_{\min}) \right\}$$

If $t^\star_n(p_{\min}) > -\infty$ or equivalently $t^\star_n(p_{\min}) \in D_{\mathrm{Pr}}$ we define the primal PULSE problem and its solution by (Primal $.t^\star_n(p_{\min}).n$) and $\hat{\alpha}^n_{\mathrm{Pr}}(t^\star_n(p_{\min}))$.

Theorem 3.1 (Primal REPRESENTATION OF PULSE)

Let $p_{\min} \in (0,1)$ and Assumptions 3.3 and 3.4 hold and assume that $t^\star_n(p_{\min}) > -\infty$. If $T_n\left(\hat{\alpha}^n_{P^n_r}(t^\star_n(p_{\min}))\right) \leq Q_{\chi^2_q}(1-p_{\min})$, then the PULSE problem has a unique solution given by the primal PULSE solution. That is, $\hat{\alpha}^n_{\mathrm{PULSE}}(p_{\min}) = \hat{\alpha}^n_{\mathrm{Pr}}(t^\star_n(p_{\min}))$.

Dual representation of PULSE

Dual representation allows for the construction of a binary search algorithm for the PULSE estimator.
Define the dual problem (Dual. $\lambda.n$) by

$$\text{minimize } l^n_{\mathrm{OLS}}(\alpha) + \lambda l^n_{\mathrm{IV}}(\alpha).$$

Whenever Assumption 3.3 (a) holds, i.e., $\mathbf{Z}^\top \mathbf{Z}$ is of full rank, then for any $\lambda \geq 0$ the solution to (Dual. $\lambda.n$) coincides with the K-class estimator with $\kappa = \lambda/(1+\lambda) \in [0,1)$:

$$\hat{\alpha}^n_{\mathrm{K}}(\kappa) = \left(\mathbf{Z}^\top (\mathbf{I} + \lambda P_{\mathbf{A}}) \mathbf{Z}\right)^{-1} \mathbf{Z}^\top (\mathbf{I} + \lambda P_{\mathbf{A}}) \mathbf{Y}.$$

The PULSE estimator $\hat{\alpha}^n_{\text{PULSE}}\ (p_{\min}\ )$ solves a K-class problem $(\text{Dual}.\lambda, n)$. Define the dual analogue of the primal PULSE constraint $t^\star_n\,(p_{\min})$ as

$$\lambda^\star_n\,(p_{\min}) := \inf\left\{\lambda \geq 0 :\ T_n\left(\hat{\alpha}^n_K(\lambda)\right) \leq Q_{\chi^2_q}\left(1 - p_{\min}\right)\right\}.$$

If $\lambda^\star_n\,(p_{\min}) < \infty$, we define the dual PULSE problem by (Dual. $\lambda^\star_n\,(p_{\min})\,.n$) with solution $\hat{\alpha}^n_{\mathrm{K}}\,(\lambda^\star_n(p)) = \arg\min_{\alpha\in\mathbb{R}^{d_1+q_1}} l^n_{\mathrm{OLS}}(\alpha) + \lambda^\star_n\,(p_{\min})\, l^n_{\mathrm{IV}}(\alpha)$.

Theorem 3.2 (Dual REPRESENTATION OF PULSE)

Let $p_{\min} \in (0, 1)$ and Assumptions 3.3, 3.4, and 3.6 hold. If $\lambda^\star_n\,(p_{\min}) < \infty$, then it holds that $t^\star_n\,(p_{\min}) > -\infty$ and $\hat{\alpha}^n_{\mathrm{K}}\,(\lambda^\star_n\,(p_{\min})) = \hat{\alpha}^n_{\mathrm{Pr}}\,(t^\star_n\,(p_{\min})) = \hat{\alpha}^n_{\mathrm{PULSE}}\,(p_{\min})$.

Simulation and Experiments Omitted.
Thank you.